UNIDAD 3: Práctica 14 - Distribuciones de probabilidad discreta

Abigail Ramos

1/9/2022

INTRODUCCIÓN A LAS DISTRIBUCIONES DE PROBABILIDAD.

La teoría de la probabilidad y de variable aleatoria van a permitir establecer un amplio catálogo de modelos teóricos, tanto discretos como continuos, con los cuales se van a poder asimilar muchas de las situaciones de la vida real. El estudio de los modelos teóricos, incluyendo la caracterización a través de sus parámetros, el cálculo de probabilidades en sus distintos formatos y la generación de números aleatorios, van a facilitar enormemente el análisis de estas situaciones reales, algunos ejemplos de estos fenómenos son:

- Si se contesta al azar un examen tipo test de 10 preguntas, donde cada una de ellas tiene 4 posibilidades siendo sólo una de ellas la correcta, ¿qué número de aciertos es más probable?
- Se sabe que las bombillas de bajo consumo de 14 w tienen una vida media útil de 10,000 horas, mientras que las bombillas clásicas por incandescencia de 60 w tienen una vida media útil de 1,000 horas. Si cada día se encienden unas 4 horas ¿cuál es la probabilidad de que después de un año estén funcionando las dos?, ¿ninguna de las dos?, ¿al menos una de las dos?

El primer problema a resolver será la elección del modelo teórico apropiado para cada caso en estudio. Para tener un buen manejo matemático de las distintas situaciones que se puedan plantear dada la distinta naturaleza y la diversidad de los resultados que proporcionan los experimentos, se necesita realizar una abstracción cuantificada del experimento. Esto lleva a una primera gran clasificación entre modelos de probabilidad discretos y continuos.

Las probabilidades asociadas a cada uno de los valores de la variable aleatoria pueden ser organizadas como una distribución de probabilidad, expresándose mediante una tabla, una gráfica o una fórmula, denominándose en este último caso, a la regla de correspondencia valores – probabilidades, función de probabilidad.

Como sabemos, los números aleatorios son descritos por una distribución. Esto es, alguna función la cual especifica la probabilidad que un número aleatorio este en algún rango, por ejemplo $P(a < X < b)$. Frecuentemente es dada por una densidad de probabilidad (en el caso continuo) o por una función masa de probabilidad PX x $P(X = x) = p(x)$ en el caso discreto. Con R podemos obtener números seleccionados aleatoriamente de diferentes distribuciones, para ello sólo tenemos que familiarizarnos con los parámetros que hay que dar a las funciones tal como la media, o una proporción, etc (dependiendo de la distribución que se esté considerando y de lo que se esté analizando).

DISTRIBUCIONES DISCRETAS

CÁLCULO DE PROBABILIDADES.

- **Ejemplo 1:** Si un estudiante responde al azar a un examen de 8 preguntas de verdadero o falso.

a) ¿Cuál es la probabilidad de que acierte 4? $P[X = 4]$ La variable X="número de aciertos" sigue una distribución Binomial de parámetros $n = 8$ y $p = 1/2$ (p probabilidad de éxito).

Para $P[X=4] = \binom{8}{4} \cdot (0.5)^4 \cdot (0.5)^{8-4} = 0.2734375$

```
#usando la funciones propias de R
dbinom(4,8,0.5)
```

```
## [1] 0.2734375
```

```
# dbinom calcula la probabilidad en un valor concreto
```

b) ¿Cuál es la probabilidad de que acierte a lo sumo 2? $P[X \leq 2]$

```
x<- 2
n<-8
p<-1/2
pbinom(x, size = n, prob = p, lower.tail=TRUE)
```

```
## [1] 0.1445313
```

```
# pbinom es la función de distribución acumulada
```

c) ¿Cuál es la probabilidad de que acierte 5 o más? $P[X \geq 5]$

```
x <- 4
n=8
p=1/2

#primera forma
F <- 1 - pbinom(x, n, p, lower.tail=TRUE)
F
```

```
## [1] 0.3632813
```

```
#segunda forma
pbinom(4, size=8, prob=0.5, lower.tail=FALSE)
```

```
## [1] 0.3632813
```

- **Ejemplo 2:** Una cierta área de Estados Unidos es afectada, en promedio, por 6 huracanes al año. Encuentre la probabilidad de que en un determinado año esta área sea afectada por:

a) Menos de 4 huracanes.$P[X < 4] = F(3)$ Se define la variable X = "número de huracanes por año" y asumiendo que dicha variable tiene una distribución de Poisson de parámetro $\lambda = 6$, porque describe el número de éxitos por unidad de tiempo y porque son independientes del tiempo desde el último evento. Se calcularán ahora las probabilidades:

```
x <- 3
mu <- 6
ppois(x, lambda = mu, lower.tail=TRUE)
```

```
## [1] 0.1512039
```

b) Entre 6 y 8 huracanes. $P[6 \leq X \leq 8] = P[X \leq 8] - P[X \leq 5] = F(8) - F(5)$ Para calcular la probabilidad de que ocurran entre 6 y 8 huracanes, se pueden sumar las probabilidades $P(X = 6) + P(X = 7) + P(X = 8)$

```
#primera forma
sum(dpois(c(6,7,8),lambda = 6))
```

```
## [1] 0.4015579
```

O restar las probabilidades acumuladas, con la opción Cola izquierda, $P(X \leq 8) - P(X \leq 5)$ Como antes se realizan en primer lugar las probabilidades acumuladas y se restan los resultados obtenidos:

```
# segunda forma
F8 <- ppois(8, lambda = 6, lower.tail=TRUE)
F5 <- ppois(5,lambda = 6, lower.tail=TRUE)
F8 - F5
```

```
## [1] 0.4015579
```

c) Represente gráficamente la función de probabilidad de la variable aleatoria X que mide el número de huracanes por año.

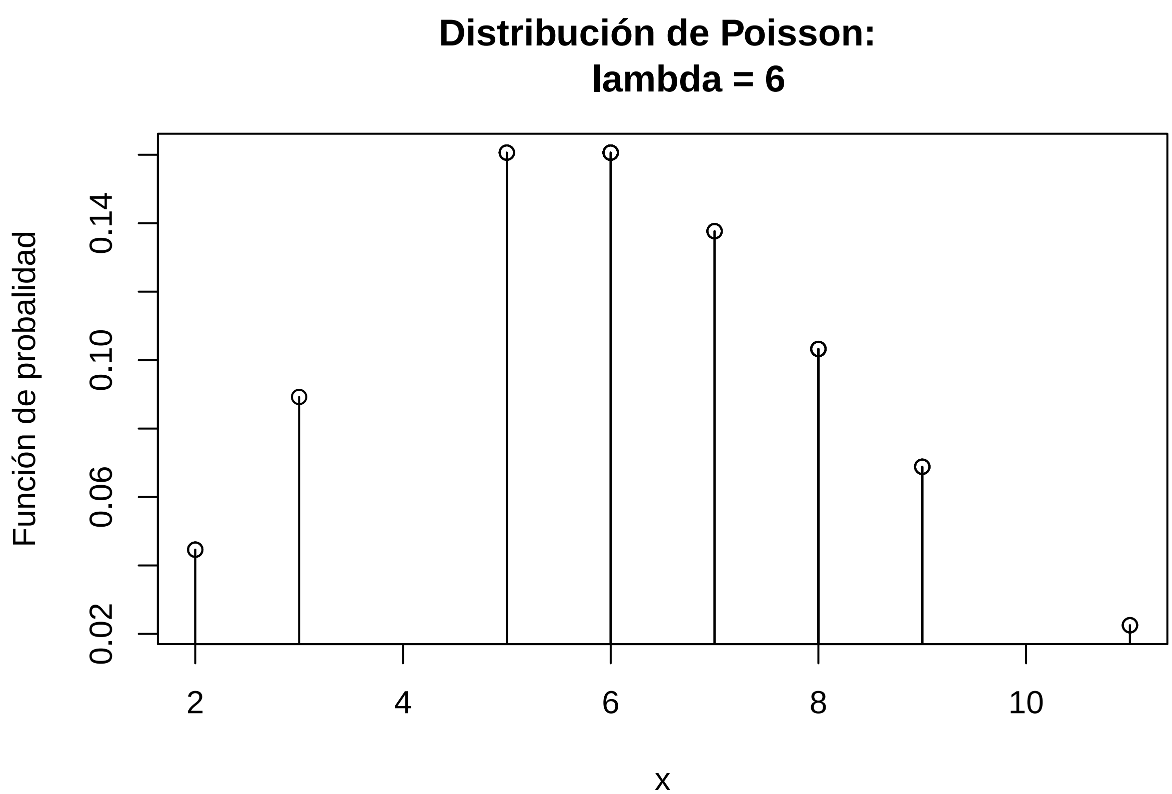
```
n <- 30

#genera 30 valores de una distribución de Poisson con lambda = 6
x <- rpois(n, lambda=mu)

#calcula las probabilidades para cada valor generado
y <- dpois(x, lambda=mu)

#genera el gráfico de distribución
plot(x, y, xlab="x", ylab="Función de probalidad", main="Distribución de Poisson:
     lambda = 6",type="h")

#une los puntos a las líneas
points(x, y, pch=21)
```



- **Ejemplo 3:** En un juego se disponen 15 globos llenos de agua, de los que 4 tienen premio. Los participantes en el juego, con los ojos vendados, golpean los globos con un bate por orden hasta que cada uno consigue romper 2.

a) ¿Cuál es la probabilidad de que el primer participante consiga un premio?

Para el primer participante la variable X="número de premios conseguidos entre 2 posibles" sigue una distribución hipergeométrica de parámetros $m = 11, n = 4, K = 2$.

```
x <- 0:2
m = 11
n = 4
k=2
# x define el número de globos con premio

# se construye la distribución de frecuencias del número de premios
Tabla <- data.frame(Probabilidad=dhyper(x, m, n, k))
rownames(Tabla) <- c('Ningún premio', 'Solamente uno', 'Dos premios')
Tabla
```

```
##               Probabilidad
## Ningún premio   0.05714286
## Solamente uno   0.41904762
## Dos premios     0.52380952
```

b) Si el primer participante ha conseguido sólo un premio, ¿cuál es la probabilidad de que el segundo participante consiga otro? Para el segundo participante la variable seguirá una hipergeométrica de parámetros $m = 10, n = 3$ y $k = 2$

```
x = 1
m= 10
n= 3
k= 2
dhyper(x, m, n, k)
```

```
## [1] 0.3846154
```

- **Ejemplo 4:** Un vendedor de alarmas de hogar tiene éxito en una casa de cada diez que visita. Calcula:

a) La probabilidad de que en un día determinado consiga vender la primera alarma en la sexta casa que visita.

Se define la variable X="número de casas que visita antes de conseguir vender la primera alarma", que sigue una distribución Geométrica con Probabilidad de éxito= 0.1 Habría que calcular la probabilidad de que tenga 5 fracasos antes del primer éxito, obteniendo de la tabla la probabilidad $P(X = 5) = 0.059049$

```
# x define el número de intentos fallidos
x <- 0:5
p=0.1

# creando la tabla de distribución de frecuencias del número de intentos fallidos
#antes de obtener la primera venta.
Tabla <- data.frame(Probabilidad=dgeom(x, prob=p))

# nombrando las filas de la distribución de frecuencias
rownames(Tabla) <- c('Venta en el primer intento', 'Venta en el segundo intento',
                     'Venta en el tercer intento', 'Venta en el cuarto intento',
                     'Venta en el quinto intento', 'Venta en el sexto intento')
Tabla
```

```
##                             Probabilidad
## Venta en el primer intento      0.100000
## Venta en el segundo intento     0.090000
```

```
## Venta en el tercer intento      0.081000
## Venta en el cuarto intento      0.072900
## Venta en el quinto intento      0.065610
## Venta en el sexto intento       0.059049
```

b) La probabilidad de que no venda ninguna después de siete viviendas visitadas.

La variable X="número de alarmas vendidas en 7 viviendas" sigue una distribución Binomial con $n = 7$ Ensayos binomiales y Probabilidad de éxito $p = 0.1$, luego en nuestro caso se tiene $P(X = 0) = 0.4782969$

```
x=0
n=7
p=0.1
dbinom(x, n, p, log = FALSE)
```

```
## [1] 0.4782969
```

c) Si se plantea vender tres alarmas, ¿cuál es la probabilidad de que consiga su objetivo en la octava vivienda que visita?

Para abordar esta cuestión, se define la variable Y= "número de casas que visita antes de conseguir vender la tercera alarma". Esta variable sigue una distribución Binomial Negativa de parámetros Número de éxitos= 3, Probabilidad de éxito $p = 0.1$, de donde: $P(Y = 5) = 0.01240029$

```
y <- 0:5
r=3
p = 0.1
Tabla <- data.frame(Probabilidad=dnbinom(y, size=r, prob=p))
rownames(Tabla) <- 0:5
Tabla
```

```
##   Probabilidad
## 0   0.00100000
## 1   0.00270000
## 2   0.00486000
## 3   0.00729000
## 4   0.00984150
## 5   0.01240029
```

GENERACIÓN DE MUESTRAS ALEATORIAS DE LAS DISTRIBUCIONES

- **Ejemplo 1:** Generar 100 números aleatorios de una distribución Binomial de parámetros n= 15 ensayos o pruebas y una probabilidad de éxito de 0.25.

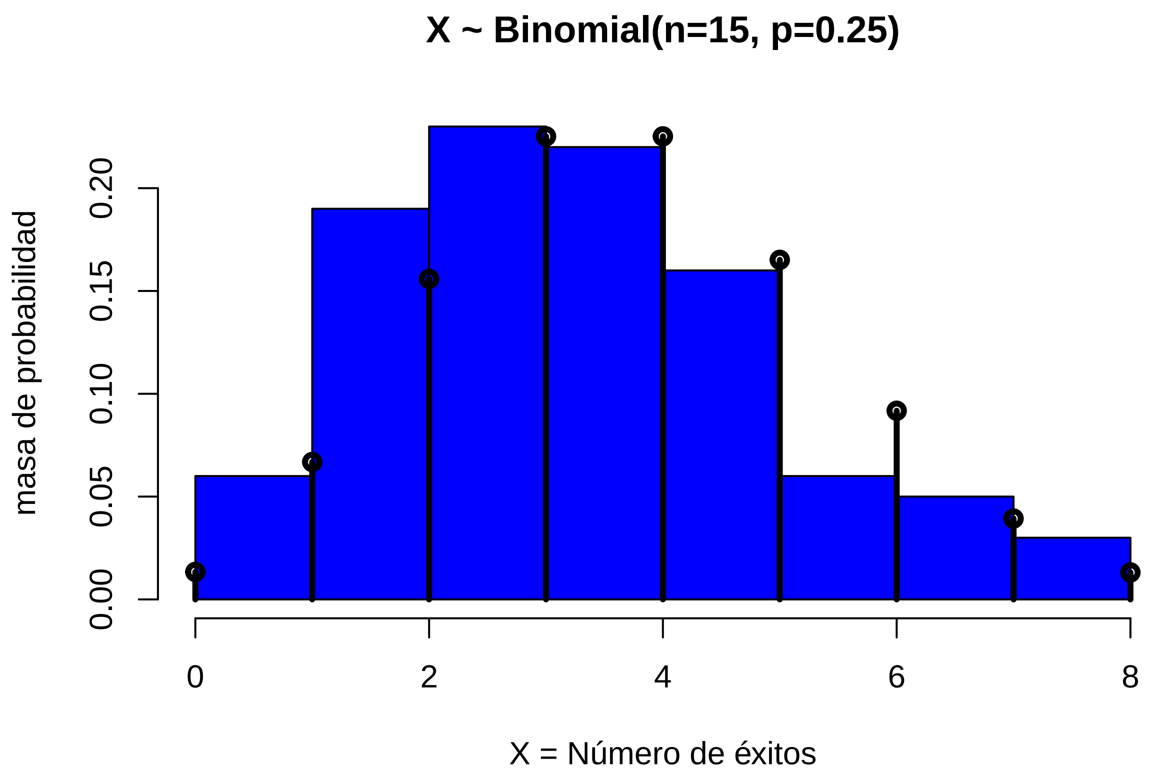
```
# Definir los parámetros apropiados
n <- 15
p <- 0.25

# generar 100 números aleatorios binomiales
x = rbinom(100, n, p)
x
```

```
##   [1] 8 3 6 2 2 5 2 4 2 1 3 2 6 3 2 8 5 3 7 4 3 3 6 4 3 4 4 2 3 6 4 3 5 1 4 2 2
##  [38] 5 0 4 2 5 3 3 3 6 5 1 6 4 2 7 2 5 3 5 3 4 5 1 4 4 2 5 4 4 5 3 2 3 5 4 3 3
##  [75] 7 4 4 3 3 5 4 5 3 4 5 4 8 3 7 2 4 2 7 2 2 1 2 4 3 5
```

```r
# Histograma para la muestra aleatoria de tamaño 100
hist(x, main="X ~ Binomial(n=15, p=0.25)", xlab="X = Número de éxitos",
     ylab="masa de probabilidad", probability=TRUE, col="blue")

# Graficar la función de probabilidad teórica, use la función points(), no debe
#cerrar el gráfico obtenido con la instrucción anterior
xvals=0:n; points(xvals, dbinom(xvals, n, p), type="h", lwd=3)
points(xvals, dbinom(xvals, n, p), type="p", lwd=3)
```



- **Ejemplo 2:** Generar 100 números aleatorios de una distribución Poisson con 200000 ensayos o pruebas y una probabilidad de éxito de 3/100000

```r
# Definir los parámetros apropiados
n <- 200000
p <- 3/100000
lambda=n*p

# generar 100 números aleatorios de la distribución
x = rpois(100, lambda)
x
```

```
##   [1]  9  3  4  7  2  5 10  5  5  8 14  6  8  7  7  1  6  5  3  8  4  8  5  3 11
##  [26]  9  4  7  6  4  3  6  9  5  5  5  3  5  9  5  2  8  7  6  5  7  5  6  7  7
##  [51]  7 10  1  3  6  6 10  2  5  4  6  2  4  8 11 12  6  6  5  2  4  6  9  5 11
##  [76]  5 10  6 11  5  6  7  9  4  5  4  4  8  6  4  7  7  5  9  6  3  7  9  3  7
```

```r
# Histograma para la muestra aleatoria de tamaño 100
hist(x, main=expression(paste("X ~ Poisson( ", lambda, " = 6 )")),
     xlab="X = Número de eventos a una tasa constante", ylab="masa de probabilidad",
     probability=TRUE, col="blue")

# Graficar la función de probabilidad teórica, use la función points()
xvals=0:n; points(xvals, dpois(xvals, lambda), type="h", lwd=3)
points(xvals, dpois(xvals, lambda), type="p", lwd=3)
```